BEBR

FACULTY WORKING PAPER NO. 1028

A Laboratory Comparison of Two Methods of Optimal New Product Concept Generation: Toward Validation

D. Sudharshan

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1028

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

March 1984

A Laboratory Comparison of Two Methods of Optimal New Product Concept Generation: Toward Validation

D. Sudharshan, Assistant Professor
Department of Business Administration

Abstract

While several analytical methods for optimal new product concept generation have been reported in the literature, no empirical comparison of their performance has been documented. This paper reports a laboratory comparison of two such methods. The results reported here add further validity to a computer simulation based comparison.

INTRODUCTION

This paper reports on a comparison of two critical modelling assumptions in the generation of optimal new product concepts. While several algorithms have been proposed in the analytical marketing literature, so far the only reported comparisons have been performed using computer simulations. Here, we report on a laboratory simulation study in which optimal new product concepts as generated by two different methods were constructed and compared. The criteria for comparisons were both predicted preference shares and actual preference shares in a sample.

Development of optimal new product concept generating methods is attracting increasing attention for the purposes of better directing marketing planning and strategy. In 1974, Shocker and Srinivasan proposed an analytic framework for the generation of optimal new product ideas. Since then several optimization algorithms have been proposed in the analytical marketing literature to generate optimal new product concepts.

These methods (algorithms) are: GRID SEARCH (suggested for this problem by Shocker and Srinivasan (1974)); PROPOSAS, Albers and Brockhoff (1977)--this method is now called PROPOPP (Albers and Brockhoff (1982)); ZIPMAP due to Zufryden (1977); the methods due to Gavish, Horsky and Srikanth (GHS) (1983), and PRODSRCH due to May, Shocker, and Sudharshan (1982), Sudharshan (1982). Each algorithm tries to find the optima of an objective function (depicting incremental preference share or incremental revenue) which is derived based on a model of consumer preferences. A general formulation of the problem is provided in Appendix A.

One of the major differences between the methods compared here, lies in the preference model incorporated in the respective methods. All the methods except GRID SEARCH and PRODSRCH, assume that it is sufficient to allow only a single choice model (k = 1, in the notation used in Appendix A), i.e., it is assumed to be enough to treat each consumer or segment as choosing the same product at all times from an assortment of products. GRID SEARCH and PRODSRCH are more flexible in permitting models which allow consumer preferences to be computed probabilistically (k > 1). The model assumes that over many occasions it is possible that a consumer could choose several products and that all his chosen products come from a subset of the products in the market called his consideration set (the size of which set is denoted by the parameter k).

Which of these methods leads to better new products? Answering this question is tantamount to answering the question of which model-algorithm combination best represents reality and solves the problem best. In a computer simulation, it is possible to define reality and test the alternate methods' performances, knowing reality. May, Shocker, and Sudharshan (1982), Sudharshan (1982) report such a comparison of the different methods, in which it was shown that methods that allowed probabilistic choice measures (k > 1), performed better (their new product concepts obtained better numerical preference shares) than those methods that did not allow such measures in market situations where in reality consumer preferences were defined to be probabilistic. While this was an important result, several questions were left unanswered. If one were to perform such comparisons in real markets, would such

findings hold? In real markets would the different product concepts result in "perceptually" different products?

As a first step toward answering these questions we performed a laboratory study, in which we developed physical realizations of the product concepts generated by two methods and compared their relative performances. We chose the two methods that May, Shocker, Sudharshan (1982) and Sudharshan (1982) found to do best in the single choice case (Gavish, Horsky and Srikanth) and the probabilistic case (PRODSRCH).

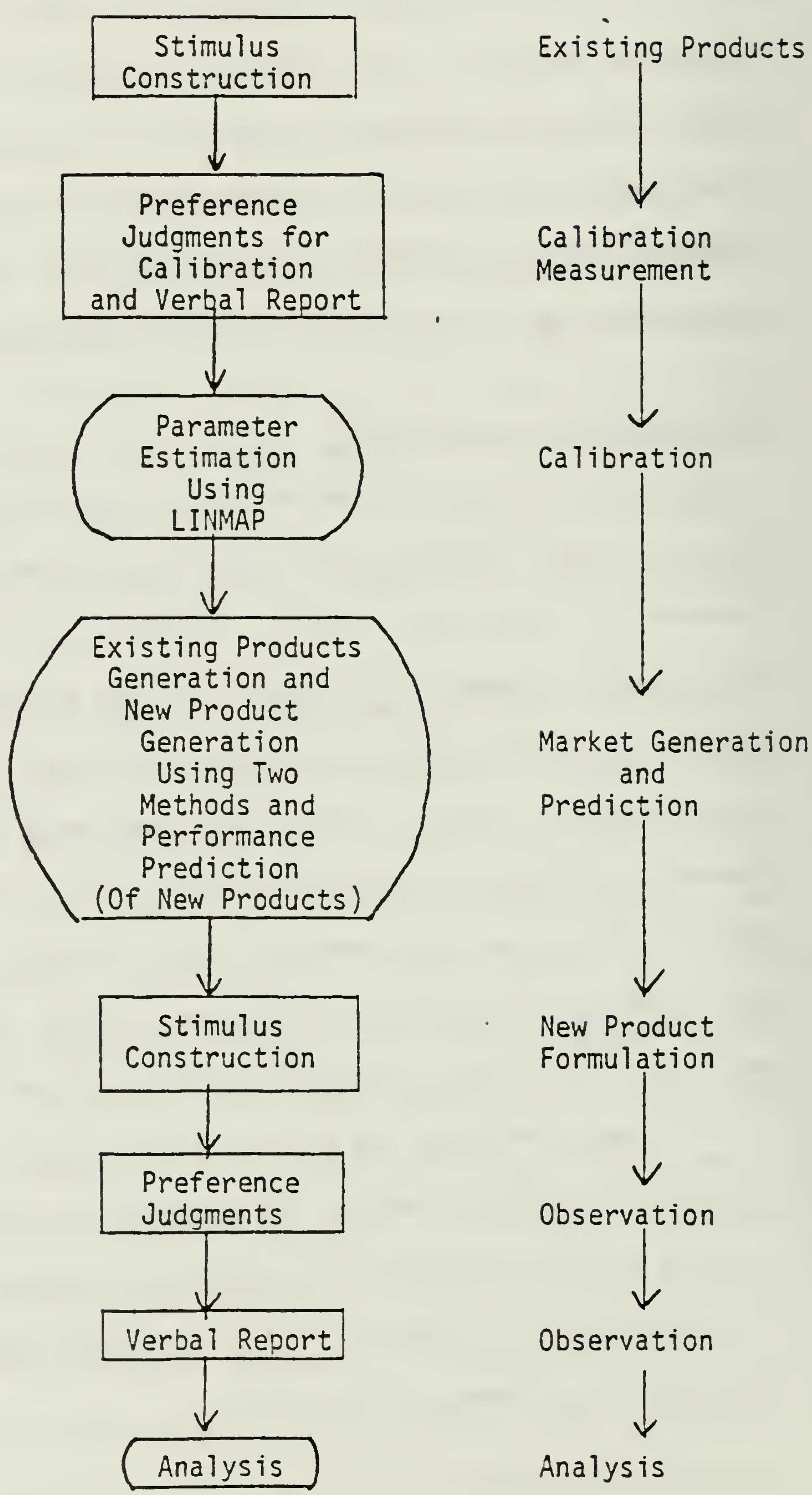
The Laboratory Study Method

In this study we carried out all the steps of the market characteristics based framework for generating optimal new product concepts (Shocker and Srinivasan (1974)). In brief, we chose a product market and obtained consumer preferences for, and perceptions of, existing products, and then generated optimal new product concepts using two search methods. These concepts were translated into new products and consumer preferences obtained for them also. This permitted us to know that we could create new products from concepts generated by PRODSRCH, and to compare the relative performance of the two "new products" to each other and to their respective predicted performances.

A schematic showing the general flow of this study leading from calibrating preference models for consumers, constructing a market of existing products, generating "optimal" new products, obtaining consumer preferences for the products in the simulated market, through analyses of the data is presented in Exhibit 1.

Exhibit 1

Scheme of Laboratory Study



As we had to formulate new products from new product concepts gathered during the study, we would not know them a priori. We had to, therefore, use a product type which could be formulated in the laboratory.

The translation of concepts gathered from the analysis into new products requires a transformation from the perceptual space of the concepts to the physical space of the products. That is, for a given concept in perceptual space, it might be possible to find more than one location in physical characteristics space. To avoid this problem we chose product characteristics for which psychophysical transformations have been generally standardized.

The product type we chose was a non-carbonated beverage, with two flavors--raspberry and lemon. Different existing products were formulated using different combinations of intensities of the two flavors. Earlier work in sensory optimization (e.g., Moskowitz (1972, 1978) using cherry flavored beverages) and in marketing (e.g., Huber (1975) using lemon-tea beverages) indicated the existence of a distribution of subjects' preferences for different flavor intensity levels.

The particular flavors chosen were such that they (a) presented a plausible combination, and (b) were not available as a combination to the marketplace (McBurney (1981)).

Also, since all the methods (exceptions being PRODSRCH and GRID SEARCH) are not designed to handle interactions among attributes, we restricted our product type to one combination of two flavors. While sugar was added to each beverage, the concentration of sugar was maintained at a constant level for all beverages to avoid possible interaction effects of sugar concentration with the different flavors.

Given the model structure incorporated in Gavish, Horsky and Srikanth's (1983) method we had to consider a product type for which we could expect finite ideal points for the different subjects. And, last but not least, for the product type we chose, we had to have reason to believe that the ideal point model, around which the different search methods have been built, would be valid for predicting preferences.

The robustness of the linear compensatory model which underlies the ideal point model has been demonstrated in the literature (e.g., Dawes and Corrigan (1974); Green and Devitta (1975); Wilkie and Pessemier (1973); and Bettman (1971, 1979)). Additionally, Huber's work (1975) demonstrated a high degree of "predictive validity" (internal) for the "ideal-point" model in predicting subjects' preferences of lemon-tea/sugar combinations using logarithm transforms of physical (concentration) space. Satisfaction of the above criteria provided us with encouragement to proceed with a "raspberry-lemon" beverage in this study.

Choice of Subjects

Subjects for this study were a convenience sample of freshman students enrolled in introductory psychology courses at a large Eastern public university. While diverse demographic groups may differ in their preference for raspberry-lemon combinations, it is less likely that a difference will be observed in the type of models that are most appropriate to modeling preference structures for each group (Huber (1975)).

We needed to have subjects who would be available for both the phases of measurement. We were able to recruit 23 subjects for the study.

While we wanted to compare the relative, observed and predicted performance of two methods of optimal new product concept generation, we were also sensitive to studying and reporting at least some aspects of consistency and validity of the data gathering and estimation methods used. We were also sensitive to various possible confounds. Our methodology was designed to reflect such concerns.

Stimulus Construction

The stimuli used for calibrating the preference models were fifteen combinations of four levels (25%, 50%, 100%, and 200%) each of raspberry and lemon beverage powders. These were chosen to represent a wide and feasible range of each attribute. Commercially available (Kool-Aid Brand) unsweetened beverage flavors were used. These percentages are relative to the manufacturer's recommended concentration (treated as 100% concentration). The sugar concentration used was also that recommended by the manufacturer. In order to avoid respondent bias caused by the color differences of the various calibration stimuli, we colored all the combinations with a food coloring to attain uniformity.

Preference Measurement

Subjects performed pairwise taste comparisons of all possible pairs of the fifteen calibration stimuli using standard taste testing procedures. To present satiation/dullness of taste buds, participants were instructed to rinse their mouths after each beverage was tasted (a pilot test was carried out with eight subjects to ensure that this procedure had the desired effects with these products). They compared one pair of beverages at a time.

Estimation of "k" (Consideration size parameter)

Verbal reports were obtained from the subjects. These reports provided information to permit estimation of the value of k for this group and also to obtain feedback regarding reactions to the study methodology.

Consistent with the laboratory study by Pessemier et. al. (1971) a common "k" was estimated for all the subjects. In that study the parameter "beta" was estimated for all subjects for a particular product class, the assumption being that "beta" was a parameter associated with product type. The parameter k (consideration set cardinality) may be expected to vary with product type and with individual consumers. It could vary with product type since each type might embrace a different number of brands, possess varied degrees of differences between brands, and the degree of involvement of consumers with products in each product type. Across individual consumers, the size of the set of products considered for purchase could also vary. When aggregation of consumers into homogeneous segments is done, perhaps our assumption of a common value of k across segments, but differing across product types, may be borne out. The question requires an empirical answer.

In the literature, the terms "evoked set," "consideration set," "awareness set," have often been (erroneously) referred to collectively as "evoked set." Consideration/evoked sets have been operationalized quite differently by various researchers. Campbell (1969), and Ostlund (1973), for example, measured "the alternatives that were considered." The phase in the decision process at which such consideration was to have taken place was not specified. Parkinson and Reilly (1979) measured those alternatives that subjects "would consider buying." Belonax and

Mittelstaldt (1978) measured the number of products that were "acceptable." And, Homans, Maddox and May (1977) measured the alternatives that were actively considered. In the stream of modeling efforts focused on in this research, the parameter k represents the set of brands that would be purchased by at least some members of a segment over a sufficiently long period of time. This would take into consideration the differences in preferences over situations and some situational exigencies such as stockouts, etc. (Shocker and Srinivasan (1977)). To estimate the value of k from observed purchase data requires that such a product be available in the market. Further, at least ten observations would have been required per subject (Blattberg, Buesing and Sen (1980)). Given our choice of product and the time frame for this project, we relied upon each subject's estimate of "the number of different fruit flavored beverages (non-alcoholic) that the subject had consumed in the last few months," as the measure for k.

This measure had the advantage that it involved only the subject's consumption. While such consumption could have occurred in the company of others, most often a range of alternative soft drinks is available for a subject to choose from based on his preference. Also, such occasions would provide a broader range of consumption situations. This measure also had the advantage of being relevant to a product type somewhat analogous to the product under study.

Calibration of Decision Models

The estimation of the parameters for the ideal point model for each subject was performed using LINMAP (Srinivasan and Shocker (1976). In particular, the strict paired comparison option in LINMAP IV was to be

used to provide increased accuracy in the estimations (see Srinivasan (1981)).

For each subject, based on the 105 pairwise preference judgments obtained, ideal point coordinates and the corresponding attribute weights were estimated. The goodness of the estimated decision model for each subject was determined based on a Kendall's tau statistic reported by LINMAP for the match between observed pairwise preferences and those predicted using the estimated model. If the estimated decision model for a subject did not lead to a significant match between observed and predicted preferences, it meant that almost any random model could have done just as well. This could happen if the subjects did not like the products considered or provided nonsense preference judgments. Care was taken to ensure that subjects understood the seriousness of the study (as part of scholarly research). The calibrations very closely corresponded with the verbal debriefings obtained from subjects at the end of the second phase of this study.

The next step was the creation of a market of existing products based on the consumer decision models thus obtained, and then to generate optimal new product concepts for such a market.

Market Generation

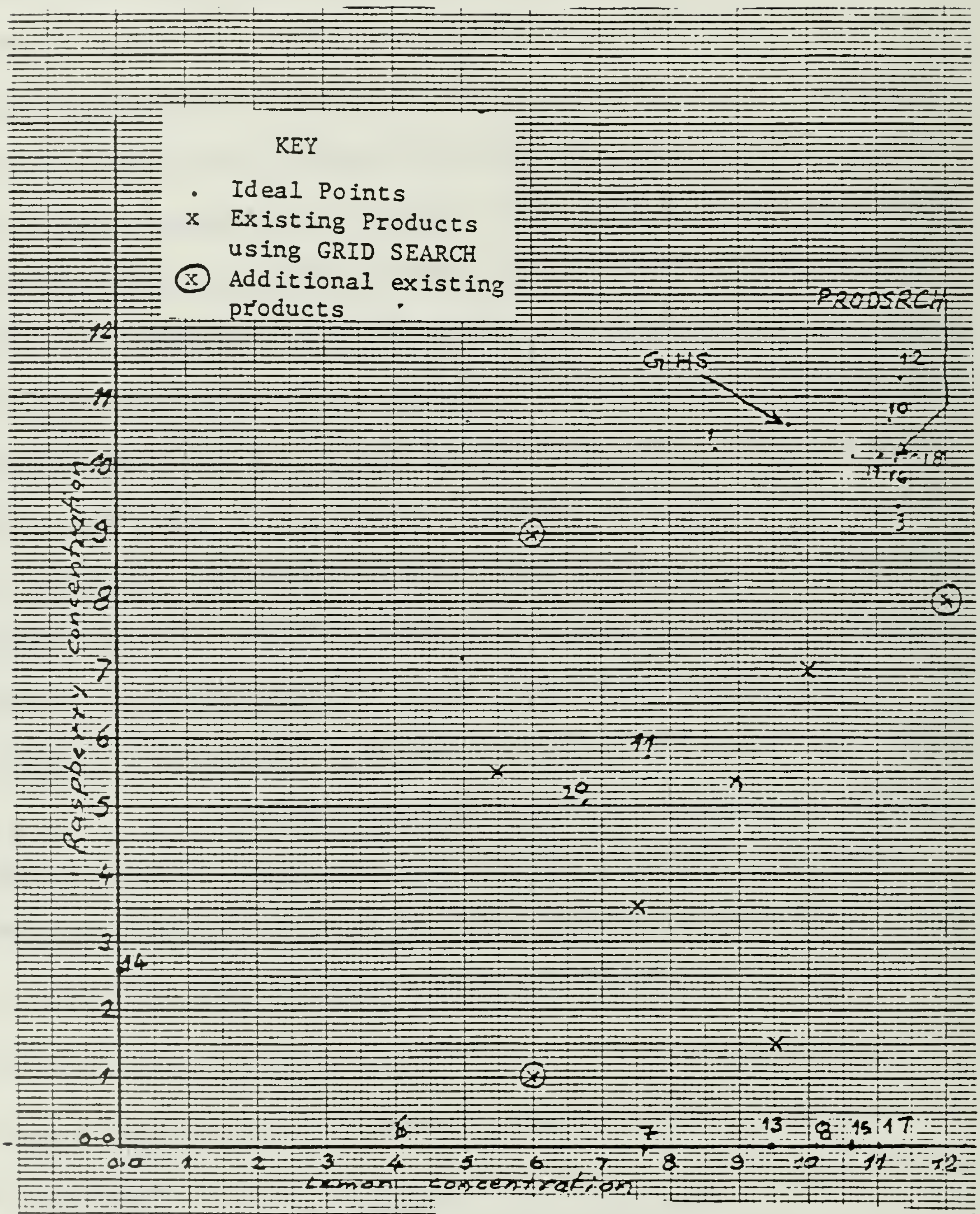
We generated existing products in the same manner as in the computer simulation comparison of May, Shocker, and Sudharshan (1982). This step avoided any bias due to subject familiarity with products used in the calibration stage. Existing products were thus generated using ideal points and attribute weights estimated earlier. A value of one for k was chosen, as all the methods compared in simulation could work

with this model. This could also be treated as a base model (e.g., Pessemier (1971); Braun and Srinivasan (1975)). We wanted to generate a set of existing products such that sufficient choice would be available for the subjects. We introduced the existing products using our sequential entry strategy with a check, the criterion being that each existing product was to be the closest product to at least one ideal point (as in the simulation) of May, Shocker and Sudharshan (1982). Five existing products were introduced using a crude grid search. To ensure validity, three more existing products were located by inspection to satisfy unfilled "demand." Subjects were, therefore, presented with ten stimuli for the second phase of measurement. Exhibit 2 is a graphical representation of the existing products, the new products, and the ideal points. As can be seen from this figure, the existing products provide a fairly wide variety of alternatives, requiring tradeoffs by subjects in their preference judgments.

Optimal new products were then generated using PRODSRCH and GHS. At this stage, the preference shares of both the new products under both $k = 1$ and $k = \hat{k}$ (the estimated value) conditions were predicted.

Product Formulation

The existing and new products generated in the previous stage had to be formulated into physical products. The product coordinates were generated in a joint space of preferences and perception. This space was described by two perceptual dimensions corresponding to flavor intensities. Each was operationalized as the natural logarithm of the concentration of a flavor.



Units (both axes):
2 x ln (concentration in gms/100U)

Exhibit 2

Final Market Configuration

Thus, to create a product we computed the concentrations of raspberry and lemon flavors that were to constitute it. This was done by obtaining inverse natural logarithms of the coordinates generated earlier. This procedure was carried out for the eight existing and the two new products. These products were then formulated as in the first stage of stimulus creation.

Observations

Preferences for these ten stimuli were obtained by two different methods from each subject. First, the ten stimuli were arranged pairwise (45 pairs). Each subject (following the same procedure as in phase one) recorded the stimulus out of each pair that he preferred. The order of presentation of the stimuli was randomized to avoid order bias. After a break of 15 minutes, subjects then ranked the ten stimuli in order of preference. The procedure used follows Green and Tull (1978, p. 480). Each subject tasted each beverage in turn and put it in one of three groups (with instructions that it was not necessary for them to place an equal number in each group) labeled:

Definitely like

Neither definitely like nor dislike

Definitely dislike

Following this, the subjects took the first group and ranked the beverages in it from most liked to least liked, and similarly for the second and third groups. (Each taste of a beverage was followed by a water rinse. The procedure for tasting was the same as that used in phase one.) By means of this two stage procedure for the second method, the

full set of ten beverages (stimuli) was eventually ranked from most liked to least liked by each subject.

Verbal Report

Following the final preference measurement, subjects were asked several questions regarding their preferences for Kool-Aid, lemon and raspberry flavors. These questions were aimed at obtaining responses which could be used to provide face validity ("does it make sense") for the ideal points and attribute weights estimated by LINMAP (Srinivasan and Shocker (1973)). Subjects were provided with a description of the study they participated in and gratitude was expressed.

Analysis

In Exhibit 3 we have provided a schematic representation of the plan of analysis followed.

FACE VALIDITY (1)[1]

The standardized attribute weights estimated for each subject were compared with his verbalizations regarding the relative importance he would place on the concentrations of the two flavors (lemon and raspberry) in deciding between beverages with different combination of the two flavors. The estimated standardized attribute weights are shown in columns 2 and 3 of Table 1 for each subject. The first sixteen subjects in Table 1 were those for whom models could be estimated with significant internal validity and finite ideal points. For subjects with serial numbers 20, 21 and 23 in Table 1, preference orderings predicted on the

EXHIBIT 3

Scheme of Analyses

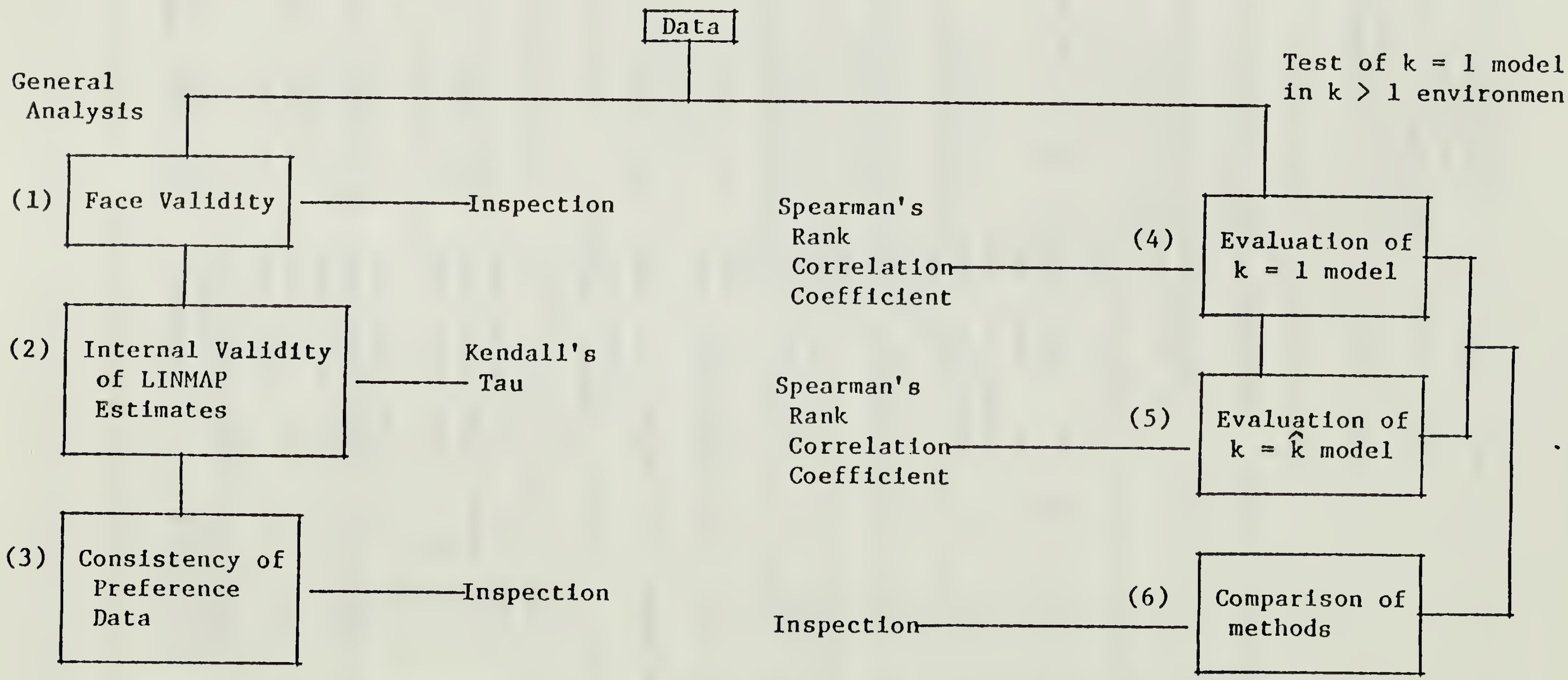


Table 1

Face Validity Tabulation

Subject Number	Att. Wt. Lem.	Att. Wt. Rasp.	Comments Attribute Weight	Ideal Point Lem.	Ideal Point Rasp.	Comments on Concentration Desired
1	0.25	0.75	Rasp. impt.	4.3	5.1	Rasp. on stronger side
2	0.68	0.32	Lemon more impt.	5.7	4.7	More sweetness
3	0.69	0.31	Lemon more impt.	2.02	0.0	More rasp.
4	0.81	0.19	Lemon more impt.	3.8	0.0	Lot of lemon Hardly any rasp.
5	0.85	0.15	Rasp. more impt.	5.05	0.0	Mod. conc.
6	0.83	0.17	Lemon more impt.	5.6	5.3	Mod. More Rasp.
7	0.67	0.33	Lemon more impt.	3.8	2.9	More rasp. than L
8	0.94	0.06	Strongly prefers lemon (doesn't like sweets)	5.7	5.65	Strong beverage
9	0.82	0.18	Lemon more impt.	4.7	0.0	A little rasp. Mod. lemon
10	0.5	0.5	Rasp. over lemon	0.0	1.3	Light conc.
11	0.95	0.05	Lemon much more impt.	5.3	0.0	Stronger lemon conc.
12	0.67	0.33	Lemon more impt.	5.6	5.02	Mod. amount of both
13	0.95	0.05	Lemon (but likes sweets)	5.5	0.0	Strong conc.
14	0.68	0.32	Lemon more impt.	5.7	5.1	Mod. and More rasp.
15	0.48	0.52	Greater weight to raspberry	5.5	5.1	Strong conc.
16	0.57	0.43	Rasp. more impt.	3.4	2.5	Mod., and More rasp.
17	0.66	0.34	Rasp. preferred	0.0	0.0	Weaker the better
18			Like lemon a lot	100.0	10.8	Like lemon a lot
19	-	1.0	Rasp. impt.	IRR	4.93	Doesn't like Kool Aid
20	0.4	0.6	Lemon more impt.	5.7	5.7	Mod., and More R the
21	0.15	0.85	Lemon more impt.	0.0	5.3	
22	0.5	0.5	Doesn't like	0.0	0.0	Doesn't like R or L
23	0.42	0.58	Not available	6.6	5.3	Not available

L - Lemon
R, Rasp. - Raspberry
Conc. - Concentration
Mod. - Moderate
IRR - Irrelevant
Impt. - Important

basis of estimated models showed a poor (insignificant) fit with the corresponding preference orderings on which the estimates were based. For subjects 17, 18, 19, and 22, it was not possible to fit significantly good models with "finite" ideal points.

In general, there is a good correspondence between respondents' claimed and estimated relative importances for the two flavors, as can be seen from the summary chart, Table 2. There were only two mismatches (out of 16) between the estimates and the verbalizations of which attribute was the more important one. A closer examination of columns 2, 3 and 4 of Table 1 shows, in general, a good correspondence between the estimated magnitudes of the attribute weights and the qualifications used in verbalizations.

For example, subject (#8) eight's estimated attribute weights for lemon and raspberry are 0.94 and 0.06 respectively and he "strongly prefers lemon"; subject (#11) eleven's estimated weights are 0.95 and 0.05 for lemon and raspberry respectively, and he considers "lemon much more important."

Examination of the estimated ideal points and corresponding verbalization for subjects (#17-22) seventeen through twenty-two (columns 5, 6, and 7; Table 1) indicates, in general, a good match between the two. In the instances of infinite ideal points, the verbalization suggested the same. For instance, subject seventeen claimed to like very weak concentrations ("the weaker the better") and the ideal estimated for her was 0.0, 0.0); for subject (#19) nineteen, one of the attributes (lemon) was estimated to be irrelevant. She claimed not to like Kool-Aid.

Table 2

Attribute Weight Face Validity Summary

		Number of Subjects Claiming to Consider This Attribute as the More Important One: Lemon	Number of Subjects Claiming to Consider This Attribute as the More Important One: Raspberry
Number of Subjects estimated to consider this attribute as the more important one	Lemon	11	2
	Raspberry	0	2

Total 16* subjects

*Importance weights were tied for one subject.

INTERNAL VALIDITY OF ESTIMATION

Table 3 shows the value of Kendall's tau corresponding to the model fitted for each subject. This statistic reflects the correlation between the preference orderings of the calibration stimuli provided by a subject and the preference ordering of the same stimuli predicted using the model fitted for the same subject. For sixteen subjects (with finite estimated ideal points) this statistic was significant at the 5% level of significance ($\alpha = 0.05$). In other words, for 88.9% of the subjects finite ideal points models with significant fits were estimated. So as not to confound our results with poor preference model estimation fit, it was decided to use internal validity as a requirement for including a subject's model for further comparisons.

CONSISTENCY OF AGGREGATE PREFERENCE MEASURED

Since the study was concerned with comparing the methods based on the preference shares observed, the concern was with the consistency of the aggregate preference measured. In other words, the stability of the aggregate preference obtained from the subjects was to be examined. Therefore, two measures of preference for the ten products (stimuli used in the second phase of measurement) were obtained for each subject. The first set of preferences was obtained using the method of "pairwise comparisons," and the second, using the method of "rank-ordering." For each pair of products (a,b) (45 pairs in all) the number of subjects preferring product "a" over "b" in "paired comparisons," and also the number of subjects preferring "a" to "b" in "rank orderings" was observed.

Table 3

Internal Validity Tabulation
of LINMAP Models

Subject St. #	Kendall's $\neq$ for model fitted
1	0.61
2	0.35
3	0.39
4	0.39
5	0.72
6	0.37
7	0.54
8	0.32
9	0.58
10	0.79
11	0.53
12	0.72
13	0.68
14	0.28
15	0.63
16	0.81
17	0.28 (both ∞)[a]
18	0.46 (lemon ∞)[a]
19	0.11 (lemon irrelevant)[a]
20	0.03 (both finite)[a]
21	0.2 (lemon ∞)[a]
22	0.52 (both at ∞)[a]
23	0.29 (both finite)[a]

a - Paranthetical comments refer to the ideal levels of the two attributes - lemon and raspberry.

An indication of the consistency of the preference share measure is provided by the discrepancies observed between the number of subjects preferring product "a" to "b" in one set of preference measurements to the number preferring "a" to "b" in the other set of preference measurements.

The mean discrepancy over the 45 pairs was 1.35, with the mode being 0, standard deviation being 3.83 and the coefficient of variation was 0.26. Thus indicating, in general, satisfactory consistency of the measure of preference share.

CHOICE OF THE MORE APPROPRIATE MODEL

To study the importance of the value of the parameter, "k," the fits between the observed ranking of the products and that predicted using each of the two models with $k = 1$, and $k = \hat{k}$ respectively were compared. Before comparing the relative performance of the new products of PRODSRCH and GHS with their relative predicted performance, establishment of the appropriate measure of relative predicted performance was required. Some search methods cannot utilize all the information available about a particular market, in their search for an "optimal" new product concept. For instance, PROPOSAS, ZIPMAP, and GHS cannot utilize information regarding the value(s) of "k." As shown in the analysis of the computer simulation data, by Shocker, May, and Sudharshan (1982), the non-utilization of this information, regarding the value of "k," could prove quite costly. For example, if $k = 1$ was actually the case, GHS would, in general, locate the product with the highest preference share (in the case of markets with equal segment sales potentials).

However, if in fact the value of k was other than 1, PRODSRCH would, in general, find a better new product concept than was found by GHS.

$\hat{k}$ was estimated by obtaining the mean of the number of different fruit flavored beverages consumed (in the last few months) by the subjects. $\hat{k}$ was thus estimated to be 4.

In this study, k = 1 was used in locating a new product with GHS, and k = 4 for generating an "optimal" new product concept for our market with PRODSRCH. (Reminder: these two methods were chosen as they were found by May, Shocker, and Sudharshan (1982) to be the best under these respective assumptions of the value of k.)

In the analysis of the laboratory data, therefore, the predicted preference shares for the new product concepts with which to compare their relative observed preference shares must first be established by evaluating the accuracy of both the k = 1 and the k = 4 predicted preference shares.

Evaluation of "k = 1" Model

Since the focus is on the relative performance of different products we examined the relationship between the predicted relative performance of all the 10 products using the k = 1 model (in the second phase of measurement) with their relative performances. The ten products we ranked on the basis of the number of subjects they were predicted to "capture" (were closest to). In Table 4, this ranking is referred to as "Predicted Ranking."

The same ten products were ranked on the basis of the information obtained from the subjects. Each subject had provided a ranking of all

Table 4

Observed Ranking of Products
(Based on Preference)

I. Consideration Sets by Subject

Subject	Products (left to right in order of decreasing preference)
1	3,8
2	4,1,5,2,3
3	1
4	9
5	2
6	4,6,3,9
7	3,6,5,7
8	6
9	1
10	5,2,6,7
11	8,9,10
12	3,9
13	7,5,4,6
14	1,6
15	7,5,4,6
16	6,2

II. Ranks by Product (within consideration sets)

Product	Rank 1	2	3	4	5
1	3	1	-	-	-
2	1	2	-	1	-
3	2	1	1	-	1
4	2	-	2	-	-
5	2	2	2	-	-
6	1	3	1	2	-
7	2	-	-	2	-
8	1	1	-	-	-
9	2	1	-	1	-
10	-	-	1	-	-

III. Consideration Set Size

Size	Frequency
1	5
2	4
3	1
4	5
5	1

Mean = 2.6
Modes = 1,5

IV. Pairwise Product Comparisons

(a >b product 'a' performs better than 'b')
(a <b product 'b' performs better than 'a')
(a ≃b cannot rank)

1 > 2								
1 > 3	2 < 3							
1 > 4	2 ≃ 4	3 > 4						
1 ≃ 5	2 < 5	3 < 5	4 < 5					
1 > 6	2 < 6	3 < 6	4 < 6	5 < 6				
1 > 7	2 < 7	3 ≃ 7	4 > 7	5 > 7	6 > 7			
1 > 8	2 > 8	3 > 8	4 > 8	5 > 8	6 > 8	7 > 8		
1 > 9	2 ≃ 9	3 > 9	4 ≃ 9	5 > 9	6 > 9	7 > 9	8 < 9	
1 > 10	2 > 10	3 > 10	4 > 10	5 > 10	6 > 10	7 > 10	8 > 10	9 > 10

V. Ranking of Products

Rank	1	2	3	4	5	6	7	8	9	10
Product	1	5	6	7	3	9	4	2	8	10

10 products. Further, each subject indicated the subset of these 10 products that he would consider consuming. The development of this ranking is shown in Table 4. The ranking was derived by comparing each pair of products, based on the rankings received by each product. Subtable V of Table 4 shows the ranking of all ten products derived by studying these pairwise comparisons. Note that we could not rank products 9, 4 and 2. However, as is evident from subtable IV, they, as a group, fit between products 3 and 10.

This observed preference based ranking of the ten products was compared with the predicted ranking. The statistic used for testing the match between the two is the Spearman's rank correlation coefficient. The value of "r_s" obtained was 0.43. Thus, the null hypotheses that the match between the two sets of ranking is no better than chance cannot be rejected at the 0.05 level (using Siegel's (1965, p. 212) suggestion for sample sizes of 10 and above).

Evaluation of "k = 4" Model

In the case of the k = 4 model also we wanted to test the null hypothesis: That, "the match between predicted and observed rankings of all the products is as expected by chance;" versus the alternative, "the match between the predicted and the observed rankings of all the products is significantly (statistically) higher than can be obtained by chance." The Spearman's rank correlation coefficient was again calculated, for the ranking of the 10 products based on the k = 4 Model, and the observed ranking (as derived in Table 4).

The value of r_s calculated is 0.625, which using Siegel's (1956, p. 212) suggestion leads to a t statistic corresponding to r_s calculated of 2.265. This suggests the rejection of the null hypothesis at the 5% level of significance implying that the match obtained using the k = 4 model between the observed and predicted rank ordering is better than can be obtained by chance alone.

Comparison of "Optimal New Products"

The Sign Test

The predicted preference share of the PRODSRCH product relative to that of the GHS product was 4.33. In other words, the PRODSRCH product is expected to be the "better" optimal new product. From Table II, Table 4 it can be seen that the PRODSRCH product was ranked first by two subjects, ranked second by one and ranked fourth by one whereas, the GHS product was included in the consideration set of only one subject, as the third ranked product (this subject's consideration set size was three). This seems to indicate that the PRODSRCH product performs better than the GHS product.

The two products (PRODSRCH's and GHS's) were compared using "The Sign Test" (Siegel, 1956, pp. 68-75). A test was carried to determine if a significant number of subjects preferred the PRODSRCH new product to the one from GHS. In both sets of preference judgments obtained in the second phase, we found thirteen subjects (out of sixteen) preferring the PRODSRCH new product to that of GHS. The sign test leads to rejection of the null hypotheses at the 0.05 level of significance (calculated probability = 0.011).

The Performance of a k = 1 Model in a k > 1 World

The k = 4 model seems to provide a better fit to the observed preferences than does the k = 1 model, thus indicating the importance of using the right value of k, and suggesting that values of k other than 1 should be incorporated in models of optimal product concept generation. This same result also indicates that the ranking of the different new products in the computer simulation obtained by May, Shocker, and Sudharshan (1982) seems to be valid. That is, the ranking that would be indicated by the computer simulation is not significantly different from that observed. Generalizability to an external population would require a different design with a much larger sample size (about 200, based on sample size requirements using a chi-squared comparison of proportions design) for the testing stage. But, what should be the sample size for calibration? This size would, perhaps, be dependent on product class and the number of segments in that market. This question is itself worth an empirical investigation.

Summary of Study

In this study, we developed models of preference (ideal-point models) for a set of subjects for different combinations of raspberry-lemon flavored beverages. Using an estimated value of k, PRODSRCH located its "optimal" new product position (assuming a set of existing products). Using a value of 1 for k, GHS located its "optimal" new product position. Numerically, these positions were quite different. Also, the predicted preference shares for these two "new products" were different, with PRODSRCH producing the "better" product. In comparisons

of the new products created corresponding to the generated "optimal" position, we found that subjects could distinguish between the two products. Further, the PRODSRCH new product performed (in terms of relative preference) better than the GHS new product, giving us more confidence in using the preference measures, computed as here, in evaluating alternative new product concepts.

IMPLICATIONS FOR MANAGEMENT DECISION MAKING

The results of the research have some significant implications for marketing planners (and strategists). In general, the research findings are important to marketing planners because, to the extent that their understanding of the factors of consideration in the generation of new product concepts is improved, they may be able to better assess the quality of the solutions obtained for their situation. The multiattribute framework used permits an understanding of not only the specific optimal new product concept generated, but also permits comprehension of the product's expected competitive environment--which segments (consumers) are expected to include this product in their consideration set; and the products with which it would specifically compete with for consumer preferences.

The influence that some of the simplifying assumptions (made by modellers) have on the "quality" of new products generated indicates that it would be appropriate for managers to give these factors more explicit consideration in the choice of a decision aid system for optimal new product concept generation. Most of the optimal new product concept generation methods currently available do not allow for values

of the consideration set parameter being other than one. Given the sensitivity of solutions to this parameter, such a facility is to be desired. New product concepts generated using the k = 1 model might result in missed opportunities. Further, marketing researchers must pay increasing attention to obtaining better measures of k. Wrong measures could again lead to new product concepts that are suboptimal.

While management already knows the importance of market segmentation, the results of this research reinforce this importance. By demonstrating the sensitivity of solution quality to a combination of "k" and segment sales potentials, these results led us to believe that while market segmentation is being performed, marketing researchers may want to incorporate the "consideration set" as an important segmentation variable. Misspecification of segments in terms of both k and/or segment sales potentials could lead to false understanding of the opportunities and threats available for new product concepts--opportunities in the form of the expected preference shares, and threats in the incorrect identification of the intensity of competition from particular existing products.

The new product concept gathering methods not only generate the optimal position, but also provide information as to the extent to which existing products would be affected by the introduction of the new product. Such measures (after proper validation) would provide valuable information for management upon which to base decisions.

Finally, given the possibility of errors in measurement of key variables such as "k" and segment sales potentials, management may wish to view results of sensitivity analysis with respect to these

key variables for their specific decision context. Such analyses can only be performed by using the more "generalizable" methods. In our framework, PRODSRCH seems to be the method that permits such a flexibility while still providing superior results.

While the laboratory study demonstrated, in a limited fashion, the possibilities of creating products from (numerical) product concepts, such a demonstration was limited to products which involved the sensory perceptions of product attributes (some examples of such products are food, cosmetics and beverages). This issue of realizing actual products from concepts is important for further research.

REFERENCES

Aaker, David and John G. Myers (1974), Advertising Management, Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Albers, Sonke and Klaus Brockhoff (1983), "PROPOPP: A Program Package for Optimal Positioning of a New Product in an Attribute Space," Journal of Marketing Research, 19 (November), 606-608.

Bachem, Achim and Hermann Simon (1980), "A Product Positioning Model with Costs and Prices," European Journal of Operational Research, 7 (August), 362-370.

Belonax, J. A. and R. A. Mittelstaedt (1978), "Evoked Set Size as a Function of Number of Choice Criteria and Information Variability," in Advances in Consumer Research, Vol. 5, ed. Keith Hunt, Ann Arbor, MI: Association for Consumer Research.

Bettman, James R. (1971), "The Structure of Consumer Choice Processes," Journal of Marketing Research, 8 (November), 465-71.

Bettman, James R. (1979), An Information Processing Theory of Consumer Choice, Addison-Wesley Publishing Company.

Blattberg, Robert C., Thomas Buesing, and Subrata K. Sen (1980), "Segment Strategies for New National Brands," Journal of Marketing, Vol. 44 (Fall), 59-67.

Braun, Michael A. and V. Srinivasan (1975), "Amount of Information as a Determinant of Consumer Behavior Toward New Products," 1975 Combined Proceedings, American Marketing Association, 373-8.

Campbell, Brian M. (1969), "The Existence of Evoked Set and Determinants of Its Magnitude in Brand Choice Behavior," Unpublished Doctoral Dissertation, Columbia University.

Dawes, Robyn M., and Bernard Corrigan (1974), "Linear Models in Decision Making," Psychological Bulletin, 81 (February), 95-106.

Garey, Michael R. and David S. Johnson (1979), Computers and Intractibility: A Guide to the Theory of NP-Completeness, San Francisco: W. H. Freeman.

Gavish, Bezalel, Dan Horsky and K. N. Srikanth (1983), "An Approach to Optimal Positioning of a New Product," Management Science, 29 (November), 1277-1297.

Green, Paul E. and Michael T. Devita (1975), "The Robustness of Linear Models Under Correlated Attribute Conditions," 1975 Combined Proceedings, American Marketing Association.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5 (September).

Green, Paul E. and D. S. Tull (1978), Research for Marketing Decisions, Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Hauser, John R. and Frank S. Koppelman (1979), "Effective Marketing Research: An Empirical Comparison of Techniques to Model Consumers' Perceptions and Preferences," in Analytic Approaches to Product and Marketing Planning, Allan D. Shocker, ed. Cambridge, MA: Marketing Science Institute.

Homans, R. E., R. E. Maddox, and F. E. May (1977), "Correlates of the Level of Information Processing for Automobiles," in Proceedings of American Institute of Decision Sciences.

Huber, Joel (1975), "Predicting Preferences on Experimental Bundles of Attributes: A Comparison of Models," Journal of Marketing Research, 12 (August), 290-7.

May, Jerrold H., Allan D. Shocker, and D. Sudharshan (1982), "A Simulation Comparison of Methods for New Product Location," working paper, College of Commerce and Business Administration, University of Illinois, Urbana-Champaign.

McBurney, Donald H. (1981), Personal Communications.

Moskowitz, Howard R. (1972), "Subjective Ideals and Sensory Optimization in Evaluating Perceptual Dimensions in Food," Journal of Applied Psychology, Vol. 56, No. 1, 60-66.

Moskowitz, Howard R. (1978), "Taste and Food Technology: Acceptability, Aesthetics, and Preference," in Edward C. Carterette and Morton P. Friedman (eds.) Handbook of Perception, Vol. VIA, NY: Academic Press.

Ostlund, Lyman E. (1973), "Evoked Set Size: Some Empirical Results," Combined Proceedings, Conference of the American Marketing Association, 226-230.

Parker, Barnett R. and V. Srinivasan (1976), "A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities," Operations Research, 24 (September-October), 991-1025.

Parkinson, T. L. and M. Reilly (1978), "An Information Processing Approach to Evoked Set Information," College of Business Administration, The Pennsylvania State University, Working Paper No. 74 (September).

Pessemier, Edgar A., Philip Burger, Richard Teach, and Douglas Tigert (1971), "Using Laboratory Preference Scales to Predict Consumer Based Preferences," Management Science, 17 (February), 371-85.

Punj, Girish N. and Richard Staelin (1978), "The Choice Process for Graduate Business Schools," Journal of Marketing Research, 15 (November), 588-598.

Shocker, Allan D. and V. Srinivasan (1974), "A Consumer Based Methodology for Identification of New Product Ideas," Management Science, 20 (February), 921-38.

Shocker, Allan D. and V. Srinivasan (1979), "Multi Attribute Approaches to Product Concept Evaluation and Generation: A Critical Review," Journal of Marketing Research, 16 (May), 159-80.

Siegel, Sidney (1956), Nonparametric Statistics for the Behavioral Sciences, NY: McGraw Hill.

Silk, A. J. (1969), "Preference and Perception Measures in New Product Development: An Exposition and Review," Industrial Management Review, (Fall), 21-37.

Silk, Alein H. and Glen L. Urban (1978), "Pre-test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15 (May), 171-191.

Srinivasan, V. (1981), "A Strict Paired Comparison Linear Programming Approach to Non-Metric Conjoint Analysis," paper presented at the Conference on Analytical Approaches to Product and Marketing Planning, Nashville, TN (October).

Sudharshan, D. (1982), "On Optimal New Product Concept Generating Methods: A Comparison of Methods," unpublished doctoral dissertation, Graduate School of Business, University of Pittsburgh.

Wilkie, William L. and Edgar A. Pessemier (1973), "Issues in Marketing's Use of Multi-Attribute Attitude Models," Journal of Marketing Research, 10 (November), 428-41.

Zufryden, Fred (1978), "ZIPMAP - A Zero-One Integer Programming Model for Market Segmentation and Product Positioning," Journal of Operational Research Society, 30 (January), 63-76.

APPENDIX A

Following Pessemier, et al. (1971), choice is modelled probabilistically as a function of this measure of preference where the individual or segment is presumed to choose from among the k-closest competitors, where k is an integer-valued parameter which can vary between 1 and the number of available brands. We operationalize this framework in terms of the following notation. Let:

B = the set of n_B existing brands which constitutes the product-market of interest, $j = 1, 2, \ldots, n_B$.

M = the set of n_M individuals and/or market segments which represent demand for the products in B, $i = 1, 2, \ldots, n_M$.

$A^{(n_A)}$ = the n_A-dimensional space spanned by determinant product attributes, i.e., $p = 1, 2, \ldots, n_A$.

$R^{(n_A)}$ = a major subspace of A in which existing and new products may feasibly be located. R is determined by technological, economic, and managerial constraints. $R \neq A$, in general.

$Y_j = \{y_{jp}\}$ = the modal perception (over all segments in M) of the j^{th} product on the p^{th} dimension in A.

$W_i = \{w_{ip}\}$ = the set of attribute weights for the i^{th} segment, reflecting the relative effect of the p^{th} attribute in the i^{th} segments' preference decision-making.

$I_i = \{I_{ip}\}$ = the most desired attribute levels ("ideal point") of the attributes for the i^{th} market segment. This ideal point will be assumed finite, but it need not lie in R.

d_{ij} = the evaluation of the j^{th} product alternative by the i^{th} market segment. This evaluation may be in the form of a preference rating, intention to buy, etc. Several alternative definitions of d_{ij} (also interpretable as a measure of proximity of the j^{th} product to the i^{th} segments' ideal point) have been proposed in the literature. The alternative models are generally special cases of the weighted Euclidean model (1) and are examples of what Green and Srinivasan (1978) have termed conjoint analysis models.

$$d_{ij} = [\sum_{p=1}^{n_A} (I_{ip} - y_{jp})^2 w_{ip}]^{1/2} \qquad (1)$$

S_i = the i^{th} segments' demand (in \$ or units) for all products in B over the period. S_i will be presumed constant.

Π_{ij} = the share if the i^{th} segments' demand allocated to the j^{th} product alternative. $\Pi_{ij} = f(d_{ij}^{-1})$ and

$$\sum_{j=1}^{n_B} \Pi_{ij} = 1 \quad \text{for all } i = 1, 2, \ldots, n_m$$

Following Bachem and Simon (1981) and Shocker and Srinivasan (1974), several forms for Π_{ij} (decision rules) can be considered:

Case 1. Every available alternative could have some non-zero likelihood of purchase, e.g., $\Pi_{ij} = a_i/d_{ij}^b$ where $a_i = 1/\sum_{j=1}^{n_B} (1/d_{ij}^b)$ and b is a parameter which varies with the product class (Pessemier, et al. 1971). Since producers would tend to locate their products at or near concentrations of demand; if ideal points are distributed throughout the space and/or attribute weights vary substantially across segments,

this decision rule should lead to relatively high likelihoods of selection for some products and low ones for others (with some arbitrary assignment of segment demand to any product located precisely at the segment's ideal point (if this occurred)). This rule says that whether or not a segment purchases a brand, there is always the potential to do so, particularly if the time period over which predictions are expected to hold is long. As a model of segment behavior, it is more credible than as a model of individual behavior, where individuals often are observed to restrict their purchases to many fewer than all available brands (Silk and Urban 1978).

Case 2. Those who argue individuals would rarely purchase brands they did not like (or judged unsuitable for their intended usage or with which they were unfamiliar), might prefer a rule which limited positive probabilities of purchase to a subset of alternatives. Individuals are also more likely to become familiar with products which better meet their objectives, due to self-interest (Aaker and Myers 1974), therefore a parameter k (possibly k_i which varies with each individual), which restricts choice to the k "closest" alternatives, would lead to a definition of $\Pi_{ij} = a_i/d_{ij}^b$ for $d_{ij} < d_i^{(k)}$, where $d_i^{(k)}$ is the distance from the i^{th} segment's ideal point to its k^{th} closest product, and $\Pi_{ij} = 0$ otherwise.

Case 3. A third rule assumes that individuals purchase only their most preferred brand, i.e., k = 1, so that $\Pi_{ij} = 1$ for that j for which $d_{ij} = d_i^{(1)}$ and $\Pi_{ij} = 0$ otherwise. The logic for this would be compelling if choice was deterministic, and all product alternatives were equally available and familiar (that is, why should individuals purchase other

than their first choice under such circumstances?). However, since likelihood of choice will typically depend upon other factors besides product characteristics (such as convenience, availability, salesperson recommendations, brand last purchased, and special situations) one would expect some variance in actual behavior. Surprisingly, then, Pessemier, et al. (1971) found that this first choice model gave good predictions in the aggregate even though it was inferior to Case 1 (above) in predicting individual-level choice. Whether analysis at the level of market segments, rather than individuals, would affect this result is not known, and should depend upon the basis for segmentation used. Additional support for a first choice model was found by Parker and Srinivasan (1976). The conditional logit model has also been used to model frequency of first choice among groups of customers (Hauser and Koppelman 1979, Punj and Staelin 1978) with good predictive results, and represents yet another alternative to those already discussed.

The form of the objective function for optimal location of a single new product concept changes with the different forms for Π_{ij}. Assume that the firm's single objective is to maximize total incremental demand, or preference share, from the new product introduction. This means that we must account for any demand for the new product which is cannibalized from the firm's existing brands. Let

Ψ_i = the set of k out of the n_B existing products closest to the i^{th} segments ideal point,

Ψ_i^* = the set of k out of the $n_B + 1$ products, existing and new, closest to that point,

χ_i = a subset of Ψ_i consisting of existing products marketed by the introducing firm, i.e., self-products,

χ_i^* = a subset of Ψ_i^* consisting of all brands (existing and new) marketed by the introducing firm,

Π_{ij} = the set of product likelihoods of purchase before new product introduction,

Π_{ij}^* = the set of product likelihoods of purchase after new product introduction,

$x = \{x_p\}$ = the new product location,

and

L = an arbitrarily large number.

Then we wish to

$$\text{Maximize} \sum_{i=1}^{n_M} u_i \left[\frac{\sum_{j \varepsilon \chi_i^*} \Pi^*_{ij}}{\sum_{j \varepsilon \Psi_i^*} \Pi^*_{ij}} - \frac{\sum_{j \varepsilon \chi_i} \Pi_{ij}}{\sum_{j \varepsilon \Psi_i} \Pi_{ij}} \right] S_i$$

subject to:

$$d_i^{(k)} (1 - u_i) \leq \left[\sum_{p=1}^{n_A} (l_{ij} - x_p)^2 w_{ij} \right]^{1/2} < d_i^{(k)} + L(1 - u_i)$$

for all $x \varepsilon R$, and $i \varepsilon M$ where u_i is zero or one depending on whether (1) or not (0) the new product is among the k closest for the i^{th} segment.

This formulation results in a nonlinear, mixed integer programming problem, involving the location of the new product and indicators as to whether it lies within the k-closest set of products for a market segment.

If we assume that every brand alternative has non-zero probability of purchase, then the quadratic constraints never become binding (i.e., we have $u_i = 1$ for all $i \varepsilon M$), and the problem reduces to an unconstrained maximization of the objective function over R. If $1 < k < n_B + 1$, then we must consider the quadratic constraints, but the Π_{ij}'s will be continuous, except when χ_i changes for a segment. This means that the derivatives of the objective function will be well-behaved almost everywhere, so that gradient-based techniques may be of value. Finally, when $k = 1$, Π_{ij} will be non-zero for only one product, so that the objective function simplifies considerably.

The major complication in this formulation is the nonlinear constraints which serve as a linkage between the location variables and the χ_i^* sets. With a weighted Euclidean distance measure, even for k - 1, the problem reduces to an integer programming problem with quadratic constraints, which is a difficult problem to solve in a reasonable amount of time. (Technically it is NP-complete. See Garey and Johnson (1979) for a thorough discussion of this topic.)